Cell, Vol. 80, 869–878, March 24, 1995, Copyright © 1995 by Cell Press

Crystal Structure and Mutational Analysis of Human Uracil-DNA Glycosylase: Structural Basis for Specificity and Catalysis

Clifford D. Mol,* Andrew S. Arvai,* Geir Slupphaug,† Bodil Kavli,† Ingrun Alseth,†‡ Hans E. Krokan,† and John A. Tainer*
*Department of Molecular Biology
The Scripps Research Institute
10666 North Torrey Pines Road
La Jolla, California 92037
†UNIGEN Center for Molecular Biology
University of Trondheim
N-7005 Trondheim
Norway

Summary

Crystal structures of the DNA repair enzyme human uracil-DNA glycosylase (UDG), combined with mutational analysis, reveal the structural basis for the specificity of the enzyme. Within the classic α/β fold of UDG, sequence-conserved residues form a positively charged, active-site groove the width of duplex DNA, at the C-terminal edge of the central four-stranded parallel β sheet. In the UDG-6-aminouracil complex, uracil binds at the base of the groove within a rigid preformed pocket that confers selectivity for uracil over other bases by shape complementarity and by main chain and Asn-204 side chain hydrogen bonds. Main chain nitrogen atoms are positioned to stabilize the oxyanion intermediate generated by His-268 acting via nucleophilic attack or general base mechanisms. Specific binding of uracil flipped out from a DNA duplex provides a structural mechanism for damaged base recognition.

Introduction

Cells constantly face the challenge of repairing damage to their genetic material that occurs under normal physiological conditions and that can be exacerbated by environmental factors (Lindahl, 1994). In contrast with damage recognition by the ATP-dependent nucleotide excision repair systems (Aboussekhra and Wood, 1994), base excision repair by DNA glycosylases requires no additional cofactors (Barnes et al., 1993; Demple and Harrison, 1994). DNA glycosylases are therefore excellent candidates for characterizing protein–DNA interactions, damage recognition, and the removal of damaged bases from DNA at the atomic level.

Uracil-DNA glycosylase (UDG; EC 3.2.2.3), also known as uracil-N glycosylase (UNG), is the first enzyme in the base excision repair pathway for removal of uracil from DNA. Uracil in DNA may result from incorporation of dUMP instead of dTMP (Tye et al., 1977) or from spontaneous deamination of cytosine in DNA due to the inherent instability of this base (Lindahl, 1994). In addition, cytosine deamination may be catalyzed by DNA-(cytosine-5)-methyltransferases, such as the HpaII methylase, which induces a high level of C→T transition mutations in double-stranded DNA (dsDNA), suggesting that mutational hot spots at CpG sites may arise from spontaneous C→T and enzymatic C→U→T deamination of 5-methylcytosine and cytosine, respectively (Shen et al., 1992). Deamination of cytosine to uracil creates a premutagenic U:G mismatch that, unless repaired, will cause a GC→AT transition mutation. Thus, Escherichia coli mutants in the *ung* gene show a 5- to 30-fold increase in the spontaneous mutation rate (Duncan and Weiss, 1982), and a 20-fold increase in spontaneous mutations is seen for similar mutants in yeast (Impellizzeri et al., 1991).

A cDNA for human UDG has been cloned, and the corresponding gene has been named *UNG* (Olsen et al., 1989). When resting human fibroblasts enter the cell cycle, transcription of the *UNG* gene is induced late in the G1 phase, resulting in an accumulation of UDG activity in the early S phase (Slupphaug et al., 1991). The *UNG* gene promoter contains an E2F element, which is involved in the cell cycle regulation of a number of genes (Nevins, 1992) and is necessary for full promoter activity (Haug et al., 1994). The *UNG* promoter lacks a TATA box and is located within a CpG island that covers the first exon (Haug et al., 1994). This, together with its ubiquitous expression, classifies the *UNG* gene as a housekeeping gene (Larsen et al., 1992).

In eukaryotic cells, most of the UDG activity is located in the nuclei and a minor fraction in the mitochondria (Wittwer and Krokan, 1985); however, both the nuclear and the mitochondrial enzymes are encoded by the same gene (Slupphaug et al., 1993). Such sharing of an enzyme from a single gene has previously been demonstrated only for the yeast genes encoding photolyase (Yasui et al., 1992) and tRNA methyltransferase (Ellis et al., 1989). Two other human cDNAs possibly encoding enzymes with UDG activity have been reported (Meyer-Siegler et al., 1991; Muller and Caradonna, 1991). However, these are unrelated both to each other and to a group of homologous UDGs, and their significance is unclear, since the *UNG*-encoded enzyme is responsible for greater than 98% of the total UDG activity in human cells (Slupphaug et al., 1995).

The structural basis for the recognition and removal of damaged bases from DNA by repair enzymes is not fully understood (Demple and Harrison, 1994; Myers and Verdine, 1994). UDG must be extraordinarily specific, being able to discriminate between uracil, which is excised, and thymine, which is a natural component of DNA and differs from uracil by a single methyl group. UDG must also be able to distinguish uracil from cytosine, which differs only in the hydrogen-bonding pattern at the C4 and N3 positions of the pyrimidine ring.

To extend structural studies of DNA repair enzymes to human enzymes and to understand the mechanism of action of DNA glycosylases further, we have determined the

‡Present address: Biotechnology Centre of Oslo, N-0316 Oslo 3, Norway.

Table 1. Statistics for the Structure Determination of Human UDG

Parameters	Native	6-(NH_2)U[a]	Thimerosal	PCMB[b]	HgUTP	$Pt(SCN)_4$	EtHgP[c]	$Hg_3(PO_4)_2$
Concentration (mM)	–	50	1	1	1	1	1	1
Soaking time (hours)	–	24	24	24	20	24	20	20
Resolution (Å)	2.0	2.3	2.9	2.2	2.6	3.0	2.6	2.8
Data completeness (%)	93	93	83	99	95	82	93	97
Number of observations	79,140	37,002	14,494	154,696	27,359	45,685	30,850	48,012
Unique reflections	14,745	10,120	4,645	12,174	7,042	4,469	6,874	5,835
R_{sym} (%)[d]	6.7	8.87	4.6	6.2	5.1	4.6	7.6	5.6
R_{iso} (%)[e]	–	–	9.7	13.9	19.0	6.7	31.2	8.3
R_{cullis}[f]	–	–	0.54	0.53	0.52	0.66	0.58	0.55
Rms f/E_{iso}[g]	–	–	1.99	2.18	2.30	1.11	1.69	1.63
Number of sites	–	1	2	1	2	1	4	3

[a] 6-(NH_2)U, 6-aminouracil.
[b] PCMB, parachloromercuribenzoate.
[c] EtHgP, ethylmercuriphosphate.
[d] R_{sym} is the unweighted R value on I between symmetry mates.
[e] $R_{iso} = \Sigma(|F_{der}|-|F_{nat}|)/\Sigma|F_{nat}|$.
[f] $R_{cullis} = \Sigma[|F_h|-(|F_{ph}|-|F_p|)]/\Sigma|F_h|$, for centric reflections.
[g] Rms f/E_{iso}, root-mean-square heavy atom F/isomorphous lack of closure error; phasing power.

atomic structures of human UDG at 2.0 Å resolution and of a UDG-inhibitor complex with 6-aminouracil at 2.3 Å resolution. This structure of a DNA glycosylase reveals a novel parallel α/β fold for duplex DNA recognition. Analysis of the active site and the activity of site-directed mutants reveal mechanisms for recognition of uracil in DNA and for cleavage of the glycosylic bond.

Results and Discussion

Structure Determination and Quality

We have determined the crystal structure of recombinant human UDG in which the N-terminal 84 amino acids of the protein encoded by the *UNG* gene have been replaced by three N-terminal residues encoded by the vector. This recombinant protein is fully active, has been expressed to high levels (Slupphaug et al., 1995), purified, and crystallized as described in the Experimental Procedures. The structure of UDG was determined by the method of multiple isomorphous replacement (MIR) with phase information from six heavy atom derivatives (see Table 1). The MIR phases to 2.6 Å resolution calculated from the six derivatives have an overall figure of merit of 0.62. These phases were improved to an overall figure of merit of 0.85 by density modification and solvent leveling with a solvent level of 35% (Zhang, 1993).

The superb quality of the modified 2.6 Å MIR map allowed over 90% of the amino acid sequence to be built into the electron density unambiguously. The remaining residues were fit to electron density maps calculated with combined MIR and model phases. The complete model consists of 223 residues, corresponding to amino acid residues 85–304 of the *UNG* gene (Figure 1), and has been refined to an R value of 0.183 at 2.0 Å resolution with excellent stereochemistry. All of the main chain atoms and nearly all of the side chain atoms are clearly visible in 2Fo – Fc electron density maps contoured at the 1σ level.

Overall Polypeptide Fold

The polypeptide topology of UDG is that of a classic α/β protein (Richardson and Richardson, 1990) that forms a compact oblate ellipsoid with approximate dimensions of 52 Å × 30 Å × 32 Å. The structure consists primarily of a central, four-stranded, all parallel β sheet surrounded on either side by a total of eight α helices and is termed a parallel doubly wound β sheet (Figures 1–2). Two helices, α1 (Glu-87 to Phe-97) and α2 (Pro-100 to His-115), precede the small β strand β1′ (Val-118 to Tyr-119) and a short helix, α3 (Thr-127 to Gln-130). The chain then enters the β sheet at β1 (Val-139 to Gly-143) and forms a long loop that leads into α4 (Pro-168 to Asp-180) and α5 (Ser-193 to Gln-198) before reentering the β sheet at β2 (Gly-199 to Asn-204) followed by β2′ (Val-209 to Arg-210), α6 (Trp-222 to Asn-236), and the third strand of the central sheet β3

```
82          α1                    α2          β1′
MEFFGESWKKHLSGEFGKPYFIKLMGFVAEERKHYTVYPPPHQV
        *   *  *                          * * P   *
126 α3          β1
FTWTQMCDIKDVKVVILGQDPYHGPNQAHGLCFSVQRPVPPPPSLEN
*  * *          * * V* **G* DPY     * * *G* * F**       * * **  *
173 α4                      α5      β2      β2′
IYKELSTDIEDFVHPGHGDLSGWAKQGVLLLNAVLTVRAHQANSHK
*   *              *  *   * *   * GV** * N   * * *          **
219         α6                  β3          α7
ERGWEQFTDAVVSWLNQNSNGLVFLLWGSYAQKKGSAIDRKRH
  *W  *     * *   *          * * ** * G   *              *
262 β4                  α8
HVLQTAHPSPLSVYRGFFGCRHFSKTNELLQKSGKKPIDWKEL
  * *    HP****     F     *F    N  L        ** *
```

Figure 1. Human UDG Amino Acid Sequence and Residue Conservation

The primary sequence of the protein was derived from the nucleotide sequence of a clone from a human placenta cDNA library (Olsen et al., 1989). The major secondary structure elements are indicated in bold and overlined. Residues that are strictly conserved among 18 known UDG gene sequences (see Experimental Procedures) are listed underneath the sequence, and highly conserved residues are marked with an asterisk.

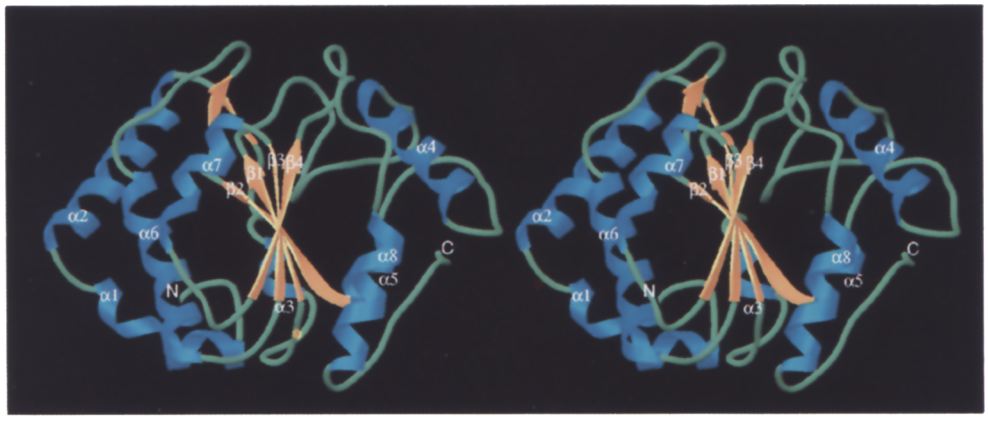


Figure 2. Overall Polypeptide Fold of Human UDG

A stereo ribbon diagram showing the secondary structure. The numbering of the β strands (yellow) and the α helices (blue) is as described in the text. The central parallel β sheet is viewed from the β4 edge with the active site of the enzyme at the top.

(Val-241 to Trp-245). The fold is completed by α7 (Ser-247 to Ala-255), the final strand of the β sheet β4 (His-262 to Thr-266), and lastly by α8 (His-283 to Lys-293). The N- and C-termini lie on opposite sides of the central β sheet, and there are no disulfide bridges in the structure. The multilayered nature of this fold gives rise to hydrophobic clusters rich in aromatic side chains that pack between the α helices and the β strands on either side of the β sheet.

A major consequence of the UDG α/β fold is a noticeable groove at one end of the β sheet (Figure 2). The width of this groove (about 21 Å diameter at the rim) is approximately equal to the diameter of a DNA double helix. Se-

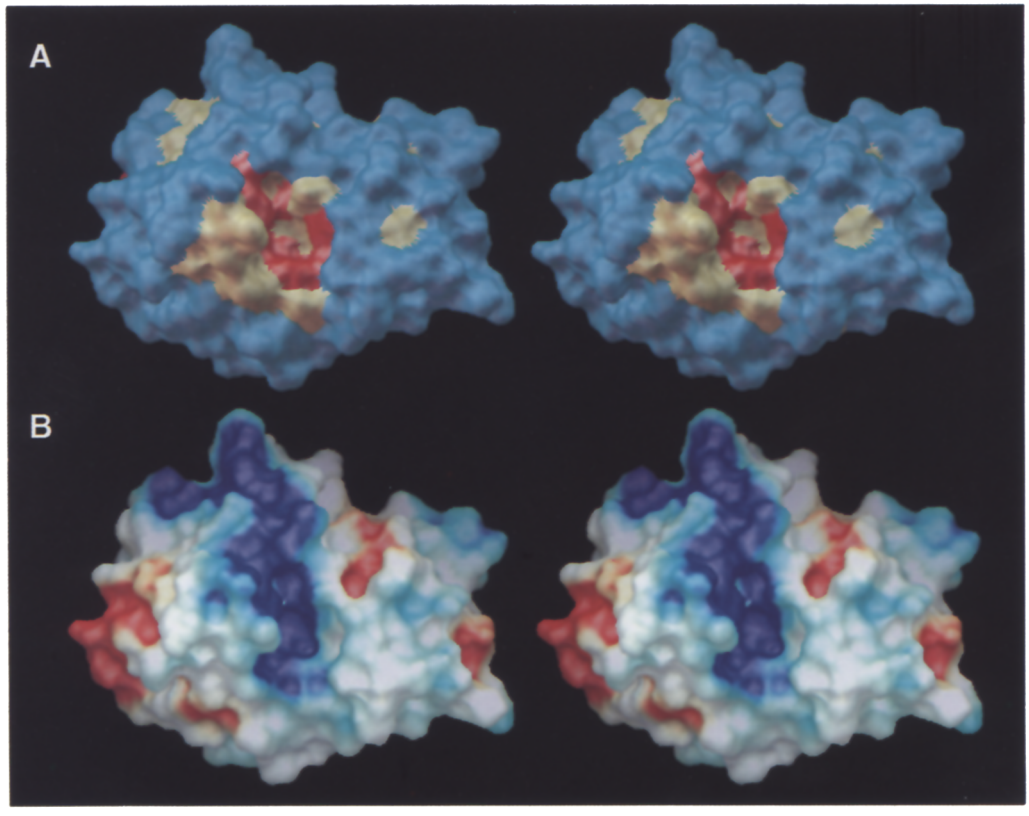


Figure 3. Molecular Surface of Human UDG Viewed Looking into the Active-Site Groove

(A) The surface is colored according to the residue conservation, with absolutely conserved residues shown in red, highly conserved residues in yellow, and nonconserved residues in blue. Conserved residues cluster in and around the active site of the enzyme at the base of the groove.
(B) The surface is colored according to the electrostatic potential with positive colored blue and negative red. Surface potential was calculated with the program DELPHI (Biosym Technologies, Inc.) using AMBER partial charges and an ionic strength of 145 mM and a radius of 1.6 Å for the solvent, with the solute and solvent dielectric constants set to 2.0 and 80.0, respectively.

Table 2. Characteristics of Site-Specific Mutants of Human UDG

Mutant	Specific Activity[a]	poly(rU) Binding[b]	DNA Binding
Wild type	100	212	166
E96Q	74	224	179
Q144L	0.15	188	153
D145N	0.04	264	180
D145E	0.08	261	173
D145E/P146G	0.11	266	163
Y147F	63	211	157
H148L	57	197	150
Q152L	98	209	177
S169A	3.9	189	172
N172V	71	211	53
W195F	51	213	161
N204Q	3.0	108	157
N204V	0.52	NB[c]	140
W245F	42	204	166
H261L	29	243	183
H268L	0.32	205	76
S270A	8.0	162	99

[a] Specific activity is expressed as a percentage of wild-type activity.
[b] poly (rU) and ssDNA binding values represent the concentration of NaCl (in mM) at which UDG eluted from the column.
[c] NB, not binding.
Specific activities and binding were measured in three to four experiments, and the mean values are listed. Binding values were determined with standard deviations $< \pm 5\%$.

quence-conserved amino acids cluster within and around this groove (Figure 3A), while basic amino acids (His-148, Arg-210, Lys-218, Arg-220, Lys-251, and Arg-276) create an area of positive electrostatic potential that would complement DNA binding along this groove (Figure 3B). Thus, the groove has the correct shape, size, and charge for UDG to bind dsDNA with the long axis of the DNA lying parallel to the groove.

The UDG α/β fold is a novel motif for protein–DNA interaction that is distinct from other known DNA repair enzymes, such as the all α-helical folds of endonuclease III (Kuo et al., 1994), endonuclease V (Morikawa et al., 1992), and the Zn loop–containing Ada protein (Moore et al., 1994). Proteins with α/β motifs similar to that of UDG (Branden, 1980) all have their active sites at the C-terminal end of the β sheet exactly where there is a groove in the UDG structure and also where the site-directed mutants that affect enzyme activity are found.

Site-Directed Mutants and Activity

Site-directed mutants of human UDG were generated by targeting residues that are conserved among the UDG sequences. Regions of the deduced amino acid sequence of the human enzyme (Olsen et al., 1989) are strikingly similar to the sequences of the corresponding proteins from E. coli (Varshney et al., 1988), Streptococcus pneumoniae (Mejean et al., 1990), Pseudomonas denitrificans (Crouzet et al., 1990), Saccharomyces cerevisiae (Percival et al., 1989), and several different herpes viruses (Baer et al., 1984; Davison and Scott, 1986; McGeoch et al., 1988; Chee et al., 1990; Worrad and Caradonna, 1988; Albrecht et al., 1992; Telford et al., 1992). In addition, UDG sequences from several members of the poxviridae are distantly, but significantly, related to the group of highly conserved UDGs (Roseman and Hruby, 1987; Goebel et al., 1990; Tartaglia et al., 1990; Upton et al., 1993; Shchelkunov et al., 1993) (see Figure 1).

Human UDG mutant enzymes were expressed and purified to apparent homogeneity, and their activities as well as both DNA and poly(rU) binding characteristics were determined (Table 2). Of the site-directed mutants characterized biochemically thus far, Asn-204 appears to be the most critical residue for the specific recognition of uracil. Even the relatively small change in the N204Q mutant decreases binding to poly(rU) by almost 2-fold, while the N204V substitution completely abolishes binding.

The mutational results (Table 2) identify four residues (Asn-204, Gln-144, Asp-145, and His-268) as most critical to UDG activity and two additional residues (Ser-169 and Ser-270) as also important. While the lack of activity in the Asn-204 mutants may reflect a loss of uracil recognition, this is not the case for the Q144L, D145E, D145N, and H268L mutants, as these mutants all retain poly(rU) binding capacity comparable to the wild-type enzyme. Thus, Gln-144, Asp-145, and His-268 appear to be involved in the reaction mechanism that cleaves the glycosylic bond.

Active Site and UDG-6-Aminouracil Complex Structures

All residues implicated by site-directed mutagenesis in the removal of uracil from DNA cluster within a groove at the C-terminal edge of the parallel β sheet. The sides of the groove are formed by 44 residues from five sequence regions: residues 144–152, 158–169, 212–218, 246–251, and 267–276. The most highly conserved residues form the sides and the floor of this groove (Figure 3A), which is implicated as the active site based upon the α/β fold, the sequence conservation, the mutagenesis results, and the binding of 6-aminouracil.

This active-site groove is long (27 Å between Ser-247 and Val-164), yet its conical shape would accommodate a DNA double helix only at its wide end. The groove narrows from about 21 Å between His-212 and Leu-272 to almost 10 Å between Pro-150 and Pro-165, suggesting that the groove floor cannot directly bind dsDNA without significant conformational change. The groove structure, however, does not suggest such conformational changes. First, the narrow end is formed by a sequence-contiguous U-shaped loop (Pro-150 to Pro-163), which on the Pro-150 side is packed against the proline-rich loop Pro-120, Pro-121, Pro-122, and on the Pro-163 side against Pro-187. Second, the two groove rims are formed by relatively tightly interacting regions. One rim is formed by the two loop regions His-212 to Asn-215 and Pro-146 to Pro-150, which have multiple interactions with other secondary structural elements. The other rim is striking in being formed primarily by an unusual seven-proline region that stretches over 22 Å, including an extended proline loop containing Pro-163, Pro-165, Pro-166, Pro-167, and Pro-168 and a second loop containing Pro-269 and Pro-271. These structural features and high local concentrations of rigid proline residues suggest that the active site does not undergo major conformational changes for binding DNA.

Polar residues in the UDG active site form hydrogen bonds with ordered water molecules, particularly within a pocket in the floor of the active-site groove (Figure 4A). Asp-145, the only acidic residue within the groove, as well as Gln-144, Asn-204, and Tyr-147 line one side of the active site. The Tyr-147 side chain OH forms a hydrogen bond with the Phe-158 backbone carbonyl oxygen, and both of these residues are completely conserved among the known UDG sequences (see Figure 1). The side chain of Phe-158 lies along the bottom of the active site, while His-268, Ser-169, and Ser-270 line the other side of the active site.

The structural basis for the exquisite specificity of UDG for uracil and for the roles of these conserved active-site residues is evident in the structure of the UDG–6-aminouracil complex. An Fo – Fc omit map (Figure 4B) of the active site of the complex illustrates how a uracil base is bound, excluding access by a uracil within a base-stacked DNA duplex. The inhibitor fits tightly within the narrow active-site recognition pocket (Figure 4C) and is encircled by residues 143–147 on three sides and by residues 157–158 on the fourth side. Three side chains play major roles in creating a recognition pocket for uracil: Asn-204, Tyr-147, and Phe-158. In the enzyme-inhibitor complex, the close approach (~3 Å) of the side chain of Tyr-147 to the base 5 position discriminates against bases substituted at this position, such as thymine or 5-bromouracil, both of which do not inhibit the enzyme (Krokan and Wittwer, 1981). The Tyr-147 ring position is fixed by van der Waals contacts with the conserved residues Val-160 and Pro-167 and by the hydrogen bond to the Phe-158 backbone carbonyl oxygen atom, creating a preexisting rigid specificity pocket. The side chain of Phe-158 lies underneath the 6-aminouracil and stacks with the pyrimidine ring (Figure 4B).

Across the active site, five hydrogen bonds between the enzyme and the uracil 2, 3, and 4 positions allow highly specific interactions of an extrahelical uracil inserted into the recognition pocket of the enzyme. The Asn-204 side chain Nδ2 and Oδ1 atoms form hydrogen bonds with the O4 and N3 positions, respectively, of the 6-aminouracil. The N3 atom of the base also forms a hydrogen bond with the backbone carbonyl oxygen of Asp-145. Both of the backbone amides of Asp-145 and Gln-144 form hydrogen bonds to the O2 atom of the base, which also forms a hydrogen bond with the Nε2 atom of His-268 via a bridging water molecule (Figure 4B). For uracil to reach its binding pocket within the UDG active site, the base must be "flipped out" from the DNA helix. Interaction with an extrahelical DNA base has been observed before in the crystal structure of HhaI (cytosine-5)-methyltransferase in complex with a 13 bp oligonucleotide (Klimasauskas et al., 1994).

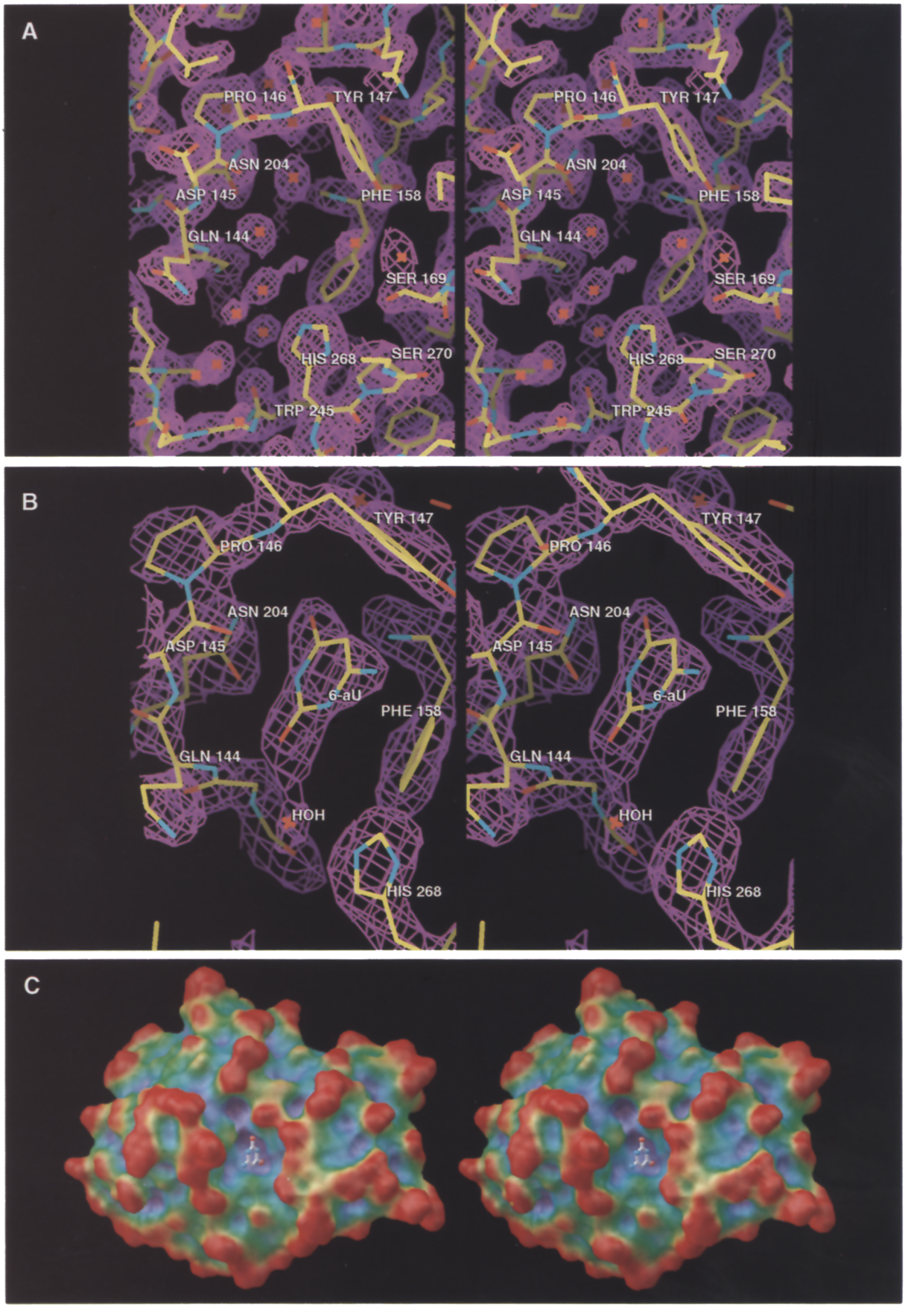
The observed binding of an extrahelical uracil is consistent with biochemical results. Uracil is removed three times as fast from single-standed DNA (ssDNA) than from dsDNA and, in general, slightly or significantly faster from U:G mismatches than from U:A mismatches (Slupphaug et al., 1995). UDG recognizes atoms of the base that would bonds. The more rapid rate of uracil removal from ssDNA than from dsDNA suggests that the displacement of uracil from the dsDNA base stack may be a rate-limiting step when the enzyme acts on dsDNA.

The active-site structure implicates specific side chains in aiding the removal of uracil from within the dsDNA base stack. On the groove rim directly above the uracil-binding pocket and protruding directly into solvent, sequence-conserved Leu-272 may be positioned to aid in locally melting the DNA helix by inserting its hydrophobic side chain between bases, as was found for a leucine in glutaminyl-tRNA synthetase complexed with tRNA (Rould et al., 1989). The side chain of the conserved residue Gln-144 is also positioned to act in initiating or stabilizing the extrahelical conformation of the uracil.

Proposed Reaction Mechanism for Glycosylic Bond Cleavage

The data obtained from mutagenesis, biochemical, and structural studies suggest two possible reaction mechanisms, both involving the imidazole group of sequence-conserved His-268. The glycosylic bond is cleaved through a two-step reaction (Figure 5). In a nucleophilic substitution mechanism, His-268 directly attacks the C1′ atom of the furanose ring to cleave the glycosylic bond. In a general base mechanism, His-268 abstracts a proton from a water molecule, creating an OH^- nucleophile that would then carry out the attack at C1′. For both mechanisms, the reaction is completed by the addition of a second water molecule, leaving a free uracil and an abasic site in DNA. The critical role of the His-268 imidazole is corroborated by the pH activity profile of the enzyme. The enzyme reaches its peak activity between pH 7.7 and 8.0, when His-268 is predominantly neutral. Near the pK_a of histidine (pH 6.5), the enzyme is only 50% active, and at pH 6.0, the enzyme is only 27% active. Yet, the enzyme loses activity very slowly as the pH is raised above 8.0 and is still 75% active at pH 10.0. As His-268 is at the surface of the enzyme and not involved in ionic interactions, it should have a pK_a near that of an isolated histidine. Although the pH activity profile of UDG is consistent with the imidazole acting either as a general base or as a nucleophile, the position of the imidazole above the uracil recognition pocket is in line to attack the C1′ atom of the furanose ring directly.

In the proposed reaction mechanism, the hydrogen bonds between the uracil 2, 3, and 4 positions and UDG Asn-204, Asp-145, and Gln-144 not only orient the nucleotide correctly but stabilize the anionic transition state upon glycosylic bond cleavage. The backbone amides of Asp-145 and Gln-144 create an oxyanion hole for the O2 atom of uracil, somewhat reminiscent of the oxyanion hole in serine proteinase structures. The hydrogen bonds of the Asp-145 backbone carbonyl and the Asn-204 side chain to the N3 and O4 atoms of the base may stabilize the oxyanion intermediate and polarize the glycosylic bond, rendering it susceptible to nucleophilic attack. Even relatively modest substitutions of Gln-144 and Asp-145 could alter these interactions, resulting in the observed catalytically impaired mutant enzymes (Table 2).


A
PRO 146
TYR 147
ASN 204
ASP 145
PHE 158
GLN 144
SER 169
HIS 268
SER 270
TRP 245
B
TYR 147
PRO 146
ASN 204
ASP 145
6-aU
PHE 158
GLN 144
HOH
HIS 268
C

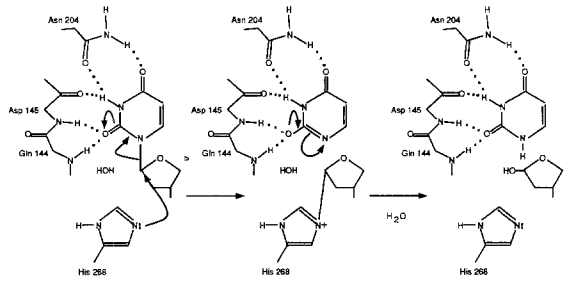


Figure 5. Schematic Representation of the Proposed Reaction Mechanism for Cleavage of the N–C1′ Glycosylic Bond by UDG

First, the nucleophilic attack by the imidazole of His-268 is facilitated by the extensive interactions made by the enzyme with the 2, 3, and 4 positions of the uracil and results in a transient covalent enzyme–DNA intermediate. The intermediate is stabilized by formation of an oxyanion at O2 that interacts with the backbone amides of Gln-144 and Asp-145. Second, the nucleophilic attack of a water molecule results in a hydrogen atom being added to N1 of uracil and an OH group added to the C1′ atom of the furanose ring. An alternative reaction mechanism would have the histidine acting as a general base and abstracting a proton from the water molecule. The N–C1′ glycosylic bond is then cleaved by a single nucleophilic attack of the resulting OH ion on the C1′ atom of the ring. Although not directly involved in the catalytic mechanism, Asn-204 interacts with the N3 and O4 positions of the base, thus ensuring that only uracil is bound at the active site. In the three-dimensional structure, the Asn-204 side chain lies deeper within the active-site groove and does not clash with the backbone carbonyl of Asp-145.

partner in the dsDNA helix and stabilize the "flipped out" orientation of the uracil. In the HhaI (cytosine-5)-methyltransferase–DNA complex, a cytosine base is swung out from the dsDNA helix, and hydrogen-bonding interactions form between a glutamine residue and the "orphan guanosine" (Klimasauskas et al., 1994). For the UDG–DNA interactions, if the uracil in DNA arose by deamination of cytosine, then the orphan base would be a guanine and the side chain Nε and Oε atoms of Gln-144 could form hydrogen bonds with the O6 and N1 atoms, respectively, of the guanine base. If the uracil in DNA arose from a U:A mispair, then only a slight rearrangement of the glutamine side chain, or the loss of one potential hydrogen bond, could still result in partial stabilization of the extrahelical uracil. The Asp-145 side chain may have a structural role by forming a pair of hydrogen bonds to the backbone amides of Asn-215 and His-217. Significant side chain rearrangements would be required for Asp-145 to participate directly in catalysis. Ser-169 Oγ and Tyr-147 OH could position the water molecule for the second stage of the reaction, while Ser-270 may interact with the DNA sugar–phosphate backbone or stabilize the His-268 imidazole orientation.

This possible reaction mechanism would also explain why UDG does not cleave the glycosylic bond when the uracil is part of a ribonucleotide rather than a deoxyribonucleotide (Talpaert-Borle et al., 1982). The RNA 2′-OH should interfere sterically with the approach of the attacking nucleophile toward C1′.

Implications for Damaged Base Recognition and Repair

Crystal structures of human UDG, combined with mutational analysis, reveal the structural basis for the specificity and activity of this critical human DNA repair enzyme that functions to prevent potentially mutagenic mispair mutations by removing uracil from DNA. UDG is a single domain α/β protein that has a distinctive active-site groove located at the C-terminal edge of a central four-stranded β sheet. Conserved amino acids cluster within and around the active-site groove, as do catalytically important residues identified by site-directed mutagenesis experiments. Structural data obtained from an enzyme-inhibitor complex clearly demonstrate that UDG must recognize uracil that is extrahelical, or "flipped out," from a dsDNA helix. UDG makes extensive interactions with the 2, 3, and 4 positions of uracil, allowing it to discriminate against cytosine, while the tight fit of the active site, particularly at the 5 position of the base, effectively discriminates against thymine or the larger purine bases. The environment and relative position of a sequence invariant and catalytically required histidine suggests a mechanism for cleavage of the glycosylic bond that involves two successive nucleophilic attacks on the C1′ atom of the furanose ring. The UDG structure thus demonstrates both a "flipped out" mechanism for identification of damaged or mismatched bases in the context of duplex DNA and the use of hydrogen bonds and steric constraints to impose specificity and selectivity. Both of these functions may be common to other DNA base excision repair enzymes.

Experimental Procedures

Materials

E. coli NR8052 (Δ(*pro–lac*),*thi*⁻, *ara*, *trp9777*, *ung1*) (Kunkel, 1985) was used for expression of wild-type and mutant proteins. E. coli JM109

Figure 4. Stereo Views of the UDG Active Site

(A) Active site electron density for the native enzyme. The 2Fo − Fc electron density map (contoured at 1.0σ) is shown with the coordinates of the uncomplexed enzyme refined at 2.0 Å resolution. The red crosses represent bound water molecules.

(B) Fo − Fc electron density omit map of the enzyme-inhibitor complex with 6-aminouracil. The inhibitor and the protein residues that interact with it were left out of the calculation of phases. The electron density map (contoured at 2.5σ) is shown with the refined coordinates of the complex, and the bound inhibitor position is well defined.

(C) Molecular surface of the UDG–6-aminouracil complex colored by shape. The 6-aminouracil (represented as a ball and stick model) is inserted into the deep pocket within the UDG active-site groove. The enzyme is colored according to its fractal density (Kuhn et al., 1992) showing that some of the most convex, protruding surface areas (red) surround the most concave parts of the active-site groove surface (blue). The shapes of the active-site groove and specificity pocket allow dsDNA binding between the groove rims, while requiring a flipped out uracil for binding the narrow pocket within the groove floor.

(*recA1, supE44, 7, e1A1endA1, hsdR17, gyrA96, r, thi, Δ(lac–proAB), F'(traD36, proAB⁺, lacIqZ M15)*) was used as host for cloning of pSELECT-1 vectors and for production of ssDNA. E. coli BMH71-18 mutS (*thi, supE, (lac–proAB), mutS::Tn10, F'(proA⁺ B⁺, lacIq M15)*) prevents repair of the newly synthesized unmethylated strand and was the first host in mutagenesis (Kramer et al., 1984; Zell and Fritz, 1987). DH5α (*supE44, lacU169(ϕ80lacZ M15), hsdR17, recA1, endA1, gyrA96, thi-1, relA1*) was the second host in mutagenesis. Expression vector pTrc99A was purchased from Pharmacia Biotechnology. The mutagenesis vector pSELECT-1 and restriction and DNA-modifying enzymes were from Promega.

Plasmid Constructions

The UNG mutants were derived from plasmid pSUNGΔ71 or from pSUNGΔ84. Mutant proteins were expressed in E. coli NR8052 after replacement of UNG restriction fragments containing the respective mutations in pTUNGΔ84 (Slupphaug et al., 1995). An EcoRI–HindIII restriction fragment (1132 bp) from pUNG15 (Olsen et al., 1989), containing the open reading frame encoding human UDG, was subcloned in pSELECT-1 and named pSUNG. Excision of an EcoRI–SacII fragment (318 bp) from pSUNG, blunting with Klenow DNA polymerase, and religation of the vector resulted in a new in-frame start codon at the ligation site. In vitro coupled transcription/translation of this construct (pSUNGΔ71) in a TNT rabbit reticulolysate system (Promega) gave a protein lacking the N-terminal 71 amino acids as compared with the full-length protein. pSUNGΔ71 was more efficiently expressed in this system, and the deleted protein had about two times higher specific enzyme activity than the full-length protein. This construct was used as a template in the mutagenesis reactions. To facilitate subcloning of 5′ region UNG mutants into pTUNGΔ84, a NcoI restriction site (CCATGG) was introduced at the start codon in pSUNG by site-directed mutagenesis. The NcoI–NcoI restriction fragment (560 bp) from pSUNG was then replaced by a NcoI–NcoI fragment (320 bp) from pTUNGΔ84, thus giving the new plasmid pSUNGΔ84 that was used as the template in mutagenesis of the 5′ region. All mutations were confirmed by DNA sequence analysis in the mutagenesis vectors (pSUNGΔ71/pSUNGΔ84) and after subcloning in the expression vector pTUNGΔ84. Site-directed mutagenesis was performed with the Altered Sites in vitro Mutagenesis System from Promega. ssDNA was prepared after infection of the JM109 strain containing pSUNGΔ71 or pSUNGΔ84 with helper phage R408. Mutant plasmids were purified using the Wizard Miniprep (Promega). Screening of specific UDG activity in mutants prior to subcloning in pTUNGΔ84 was performed by coupled transcription/translation in TNT in vitro reticulolysates as described by Slupphaug et al. (1995).

Biochemical Analysis of *UNG* Mutants

Transformed E. coli NR8052 were grown in 1 liter of LB medium with 100 μg/ml ampicillin (2.5% inoculum) at 37°C. The *trc* promoter was induced after 3 hr by addition of 1 mM IPTG (final concentration). After continued culture for 2 hr, cells were collected by centrifugation at 4000 × g for 15 min (4°C), and the bacterial pellets were frozen at −55°C. UNG mutants were purified to apparent homogeneity and verified to have the correct molecular weight by SDS–PAGE and high resolution gel filtration, as described previously (Slupphaug et al., 1995).

UDG activity was measured in 20 μl of assay mixture containing final concentrations of 10 mM NaCl, 20 mM Tris–HCl (pH 7.5), 1 mM EDTA, 1 mM DTT, 0.5 mg/ml bovine serum albumin (BSA), and 1.88 μM [^{3}H]dUMP-containing DNA (specific activity 0.5 mCi/mmol). Uracil release was measured after 10 min incubation at 30°C as described (Krokan and Wittwer, 1981). Protein concentration was measured using the Bio-Rad protein assay (Bio-Rad) with BSA as standard. For assay of general DNA binding capacity, 10 μg of each mutant was diluted to 500 μl with binding buffer (Tris–acetate [pH 7.2], 2 mM $MgCl_2$, 0.02% sodium azide, 10% glycerol [v/v]), and loaded on an ssDNA–agarose (Pharmacia) column (0.5 × 1.5 cm) preequilibrated with binding buffer. Flow was maintained at 50 μl/min using the Pharmacia SMART system equipped with a conductivity monitor and a μpeak monitor measuring absorbances at 260, 280, and 320 nm for correct identification of protein peaks. UNG mutants were eluted with a 1 ml linear gradient of NaCl (0.0–0.4 M) in binding buffer, and conductivity at peak UNG elution was recorded. For assay of uracil binding capacity, 10 μg of each mutant was diluted as above and loaded on a poly(rU)–Sepharose column (0.5 × 1.5 cm).

Amino Acid Sequence Alignment

The deduced amino acid sequences of 18 known UDG genes from different species were aligned using the Sequence Analysis Software Package, Version 7.3.1-UNIX (1991) of the Genetics Computer Group (Madison, WI). The alignment was generated using the default parameters with a gap-weight of 1.5 and a gap-length-weight of 0.1.

Data Collection and MIR Phasing

Crystals of human UDG (30 mg/ml) were grown by the hanging drop vapor diffusion technique by mixing equal volumes of protein with reservoir solutions containing 16%–20% PEG 4000 (Sigma), 100 mM imidazole/maleate (pH 7.9), and 1%–4% saturated NaCl and 1%–4% saturated ammonium sulfate. On the basis of the symmetry and systematic extinctions observed on 12° X-ray precession photographs, the crystallographic space group is orthorhombic $P2_12_12_1$ with unit cell dimensions a = 48.3 Å, b = 55.6 Å, and c = 86.3 Å, and one molecule per asymmetric unit.

X-ray diffraction data for all crystals were collected at room temperature on an Elliot GX-21 rotating anode X-ray generator operating at 40 kW and 70 mA, with a double mirror-monochromated/focused Cu Kα X-ray beam from a 0.3 mm focus and an 18 cm MAR research imaging plate. Complete data sets were obtained from single crystals using a 1° oscillation width and an exposure of 15 min per frame and were processed using MARXDS (Kabsch, 1988).

The structure of human UDG was solved by MIR using phases calculated from six heavy atom derivatives (Table 1). Heavy atom positions were determined from isomorphous difference Patterson maps and from difference and cross-difference Fourier maps calculated with the XtalView system of programs (McRee, 1992, 1993). The positions and occupancies of these heavy atom–binding sites were refined, and MIR map coefficients were calculated using PHASES (Furey and Swaminathan, 1990). This map was improved by solvent leveling and density modification with the SQUASH program (Zhang, 1993) to yield a 2.6 Å resolution map with readily interpretable electron density.

Model Building and Refinement

An almost complete model, lacking only the N-terminal nine and C-terminal five residues, a gap of four internal residues, and the side chains of six lysine and arginine residues, was built into the 2.6 Å resolution modified MIR map using the program XFIT (McRee, 1992). The remaining residues were built into a 2.6 Å resolution electron density map calculated with phases from this partial structure combined with the modified MIR phases using the program σ_A (Read, 1986).

The complete model was refined using stereochemically restrained simulated annealing and positional refinement with X-PLOR 3.1 (Brünger et al., 1987). Prior to refinement, 5% of reflections were excluded and used throughout the refinement to calculate a free R value (Brünger, 1992). A total of three macrocycles of simulated annealing with initial temperatures of 2000 K followed by 50 cycles of positional refinement were performed. Simulated-annealed omit maps were calculated with 20 residues (9%) of the model omitted from the phase calculation, and the omitted regions inspected were-rebuilt with XFIT (McRee, 1992). Further cycles of positional refinement were done to gradually include the data to 2.0 Å resolution. Refinement of individual atomic temperature values also indicated residues that needed inspection and refitting to new omit maps. Ordered solvent molecules were placed into Fo − Fc electron density maps using an automated peak picker program (Borgstahl et al., 1994), and their positions and temperature factors were refined. The free R value consistently decreased during refinement along with the conventional R value.

The native structure, containing 1,808 protein atoms and 180 ordered solvent molecules, was refined to a crystallographic R value of 0.178 for 13,503 reflections with $F > 1.0\sigma$ in the resolution range 8.0 to 2.0 Å. The R value for all data in this resolution range, 14,165 reflections, is 0.183. The free R value (Brünger, 1992) for 5% of the reflections against which the model was never refined is 0.261 for all data and 0.252 for reflections with $F > 1.0\sigma$. There are no outliers on the Ramachandran plot of the main chain torsion angles, and the error

in the atomic positions is 0.24 Å as estimated from an σ_A plot (Read, 1986). The corresponding values for the structure of the enzyme in complex with 6-aminouracil, containing 1,808 protein atoms and 152 ordered solvent molecules, are R = 0.190 for 8,304 reflections with $F > 1.0\sigma$ in the resolution range 8.0 to 2.3 Å, and free R = 0.297 for data in the same resolution range. The root-mean-square deviations of the bond distances and bond angles are 0.018 Å and 3.4° for the native structure and 0.020 Å and 3.7° for the complex structure. The average temperature factors for the native structure are 13.6 Å^2 for main chain atoms, 15.6 Å^2 for side chain atoms, and 35.9 Å^2 for solvent molecules. For the complex structure, the corresponding values are 19.1 Å^2, 20.3 Å^2, and 41.2 Å^2, with an average temperature factor for the 6-aminouracil of 17.6 Å^2.

Acknowledgments

Correspondence should be addressed to J. A. T. The authors wish to thank E. D. Getzoff, T. J. Macke, H. E. Parge, R. P. Cunningham, G. E. O. Borgstahl, C. L. Fisher, K. J. Balch, and B. R. Crane for helpful advice and discussions and S. B. Petersen for discussions related to mutagenesis. This work was supported by grants from the Norwegian Cancer Society, the Norwegian Research Council, and the National Institutes of Health grant GM 46312.

Received December 5, 1994; revised January 10, 1995.

References

Aboussekhra, A., and Wood, R. D. (1994). Repair of UV-damaged DNA by mammalian cells and *Saccharomyces cerevisiae*. Curr. Opin. Genet. Dev. *4*, 212–220.

Albrecht, J. C., Nicholas, J., Biller, D., Cameron, K. R., Biesinger, B., Newman, C., Wittmann, S., Craxton, M. A., Coleman, H., Fleckenstein, B., and Honess, R. W. (1992). Primary structure of the herpesvirus saimiri genome. J. Virol. *66*, 5047–5058.

Baer, R., Bankier, A. T., Biggin, M. D., Deininger, P. L., Farrell, P. J., Gibson, T. J., Hatfull, G., Hudson, G. S., Satchwell, S. C., Seguin, C., Tuffnell, P. S., and Barrell, B. G. (1984). DNA sequence and expression of the B95-8 Epstein–Barr virus genome. Nature *310*, 207–211.

Barnes, D. E., Lindahl, T., and Sedgwick, B. (1993). DNA repair. Curr. Opin. Cell Biol. *5*, 424–433.

Borgstahl, G. E. O., Rogers, P. H., and Arnone, A. (1994). The 1.8 Å structure of carbonmonoxy-β4 hemoglobin: analysis of a homotetramer with the R quaternary structure of liganded α2β2 hemoglobin. J. Mol. Biol. *236*, 817–830.

Branden, C. I. (1980). Relationship between structure and function of α/β proteins. Quart. Rev. Biophys. *13*, 317–338.

Brünger, A. T. (1992). Free R value: a novel statistical quantity for assessing the accuracy of crystal structures. Nature *355*, 472–475.

Brünger, A. T., Kuriyan, J., and Karplus, M. (1987). Crystallographic R factor refinement by molecular dynamics. Science *235*, 458–460.

Chee, M. S., Bankier, A. T., Beck, S., Bohni, R., Brown, C. M., Cerny, R., Horsnell, T., Hutchinson, C. A., III, Kouzarides, T., Martignetti, J. A., Preddie, E., Satchwell, S. C., Tomlinson, P., Weston, K. M., and Barrel, B. G. (1990). Analysis of the protein-coding content of the sequence of human cytomegalovirus strain AD169. Curr. Top. Microbiol. Immunol. *154*, 125–169.

Crouzet, J., Cameron, B., Cauchois, L., Rigault, S., Rouyez, M.-C., Blanche, F., Thibaut, D., and Debussche, L. (1990). Genetic and sequence analysis of an 8.7-kilobase *Pseudomonas denitrificans* fragment carrying eight genes involved in the transformation of precorrin-2 to cobyrinic acid. J. Bacteriol. *172*, 5980–5990.

Davison, A. J., and Scott, J. E. (1986). The complete DNA sequence of varicella-zoster virus. J. Gen. Virol. *67*, 1759–1816.

Demple, B., and Harrison, L. (1994). Repair of oxidative damage to DNA: enzymology and biology. Annu. Rev. Biochem. *63*, 915–948.

Duncan, B. K., and Weiss, B. (1982). Specific mutator effects of ung (uracil-DNA glycosylase) mutations in *Escherichia coli*. J. Bacteriol. *151*, 750–755.

Ellis, S. R., Hopper, A. K., and Martin, N. C. (1989). Amino-terminal extension generated from an upstream AUG codon increases the efficiency of mitochondrial import of N2, N2-dimethylguanosine-specific tRNA methyltransferases. Mol. Cell. Biol. *9*, 1611–1620.

Furey, W., and Swaminathan, S. (1990). PHASES: a program package for the processing and analysis of diffraction data from macromolecules. In ACA Meeting Summarys (New York: American Crystallographic Association), p.73.

Goebel, S. J., Johnson, G. P., Perkus, M. E., Davis, S. W., Winslow, J. P., and Paoletti, E. (1990). The complete DNA sequence of vaccinia virus. Virology *179*, 247–266.

Haug, T., Skorpen, F., Lund, H., and Krokan, H. E. (1994). The structure of the gene for human uracil-DNA glycosylase and analysis of the promoter frunction. FEBS Lett. *353*, 180–184.

Impellizzeri, K. J., Anderson, B., and Burgers, P. M. (1991). The spectrum of spontaneous mutations in a *Saccharomyces cerevisiae* uracil-DNA-glycosylase mutant limits the function of this enzyme to cytosine deamination repair. J. Bacteriol. *173*, 6807–6810.

Kabsch, W. J. (1988). Evaluation of single crystal X-ray diffraction data from a position sensitive detector. J. Appl. Crystallogr. *21*, 916–924.

Klimasauskas, S., Kumar, S., Roberts, R. J., and Cheng, X. (1994). HhaI methyltransferase flips its target base out of the DNA helix. Cell *76*, 357–369.

Kramer, B., Kramer, W., and Fritz, H. J. (1984). Different base/base mismatches are corrected with different efficiencies by the methyl-directed DNA mismatch-repair system of E. coli. Cell *38*, 879–887.

Krokan, H., and Wittwer, C. U. (1981). Uracil-DNA glycosylase from HeLa-cells: general properties, substrate specificity and effects of uracil analogs. Nucl. Acids Res. *9*, 2599–2613.

Kuhn, L. A., Siani, M. A., Pique, M. E., Fisher, C. L., Getzoff, E. D., and Tainer, J. A. (1992). The interdependence of protein surface topography and bound water molecules revealed by surface accessibility and fractal density measures. J. Mol. Biol. *228*, 13–22.

Kunkel, T. (1985). Rapid and efficient site-specific mutagenesis without phenotypic selection. Proc. Natl. Acad. Sci. USA *82*, 488–492.

Kuo, C. F., McRee, D. E., Fisher, C. L., O'Handley, S. F., Cunningham, R. P., and Tainer, J. A. (1992). Atomic structure of the DNA repair [4Fe-4S] enzyme endonuclease III. Science *258*, 434–440.

Larsen, F., Gundersen, G., Lopez, R., and Prydz, H. (1992). CpG islands as gene markers in the human genome. Genomics *13*, 1095–1107.

Lindahl, T. (1994). Instability and decay of the primary structure of DNA. Nature *362*, 709–715.

McGeoch, D. J., Dalrymple, M. A., Davison, A. J., Dolan, A., Frame, M. C., McNab, D., Perry, L. J., Scott, J. E., and Taylor, P. (1988). The complete DNA sequence of the long unique region in the genome of herpes simplex virus type 1. J. Gen. Virol. *69*, 1531–1574.

McRee, D. E. (1992). XtalView: a visual protein crystallographic software system for X11/XView. J. Mol. Graph. *10*, 44–47.

McRee, D. E. (1993). Practical Protein Crystallography (San Diego: Academic Press).

Mejean, V., Rives, I., and Claverys, J. P. (1990). Nucleotide sequence of the *Streptococcus pneumoniae ung* gene encoding uracil-DNA glycosylase. Nucl. Acids Res. *18*, 6693.

Meyer-Siegler, K., Mauro, D. J., Seal, G., Wurzer, J., deRiel, J. K., and Sirover, M. A. (1991). A human nuclear uracil DNA glycosylase is the 37-kDa subunit of glyceraldehyde-3-phosphate dehydrogenase. Proc. Natl. Acad. Sci. USA *88*, 8460–8464.

Moore, M. H., Gulbis, J. M., Dodson, E. J., Demple, B., and Moody, P. C. E. (1994). Crystal structure of a suicidal DNA repair protein: the Ada O6-methylguanine-DNA methyltransferase from *E. coli*. EMBO J. *13*, 1495–1501.

Morikawa, K., Matsumoto, O., Tsujimoto, M., Katayanagi, K., Ariyoshi, M., Doi, T., Ikehara, M., Inaoka, T., and Ohtsuka, E. (1992). X-ray structure of T4 endonuclease V: an excision repair enzyme specific for a pyrimidine dimer. Science *256*, 523–526.

Muller, S. J., and Caradonna, S. (1991). Isolation and characterization

of a human cDNA encoding uracil-DNA glycosylase. Biochim. Biophys. Acta *1088*, 197–207.

Myers, L. C., and Verdine, G. L. (1994). DNA repair proteins. Curr. Opin. Struct. Biol. *4*, 51–59.

Nevins, J. R. (1992). Transcriptional regulation: a closer look at E2F. Nature *358*, 375–376.

Olsen, L. C., Aasland, R., Krokan, H. E., and Helland, D. E. (1989). Molecular cloning of human uracil-DNA glycosylase. EMBO J. *8*, 3121–3125.

Percival, K. J., Klein, M. B., and Burgers, P. M. (1989). Molecular cloning and primary structure of the uracil-DNA-glycosylase gene from *Saccharomyces cerevisiae*. J. Biol. Chem. *264*, 2593–2598.

Read, R. J. (1986). Improved Fourier coefficients for maps using phases from partial structures with errors. Acta Crystallogr. *A42*, 140–149.

Richardson, J. S., and Richardson, D. C. (1990). Principles and patterns of protein conformation. In Prediction of Protein Structure and the Principles of Protein Conformation, G. D. Fasman, ed. (New York: Plenum Press), pp. 2–98.

Roseman, N. A., and Hruby, D. E. (1987). Nucleotide sequence and transcript organization of a region of the vaccinia virus genome which encodes a constitutively expressed gene required for DNA replication. J. Virol. *61*, 1398–1406.

Rould, M. A., Perona, J. J., Soll, D., and Steitz, T. A. (1989). Structure of an *E. coli* glutaminyl-tRNA synthetase complexed with tRNAGln and ATP at 2.8 Å resolution: implications for tRNA discrimination. Science *246*, 1135–1142.

Shchelkunov, S. N., Marennikova, S. S., Blinov, V. M., Resenchuk, S. M., Tetmenin, A. V., Chizhikov, V. E., Gutorov, V. V., Safronov, P. F., Kurmanov, R. K., and Sandakhchiev, L. S. (1993). Entire coding sequence of the variola virus. Dokl. Akad. Nauk. SSSR *328*, 629–632.

Shen, J.-C., Rideout, W. M., III, and Jones, P. A. (1992). High frequency mutagenesis by a DNA methyltransferase. Cell *71*, 1073–1080.

Slupphaug, G., Olsen, L. C., Helland, D., Aasland, R., and Krokan, H. E. (1991). Cell cyle regulation and in vitro hybrid arrest analysis of the major human uracil-DNA glycosylase. Nucl. Acids Res. *19*, 5131–5137.

Slupphaug, G., Markussen, F. H., Olsen, L. C., Aasland, R., Aarsaether, N., Bakke, O., Krokan, H. E., and Helland, D. E. (1993). Nuclear and mitochondrial forms of human uracil-DNA glycosylase are encoded by the same gene. Nucl. Acids Res. *21*, 2579–2584.

Slupphaug, G., Eftedal, I., Kavli, B., Bharati, S., Helle, N. M., Haug, T., Levine, D. W., and Krokan, H. E. (1995). Properties of a recombinant human uracil-DNA glycosylase from the *UNG*-gene and evidence that the *UNG*-gene encodes the major uracil-DNA glycosylase. Biochemistry *34*, 128–138.

Talpaert-Borle, M., Campagnari, F., and Creissen, D. M. (1982). Properties of purified uracil-DNA glycosylase from calf thymus. J. Biol. Chem. *257*, 1208–1214.

Tartaglia, J., Winslow, J., Goebel, S., Johnson, G. P., Taylor, J., and Paoletti, E. (1990). Nucleotide sequence analysis of a 10.5 kbp HindIII fragment of fowlpox virus: relatedness to the central portion of the vaccinia virus HindIII D region. J. Gen. Virol. *71*, 1517–1524.

Telford, E. A. R., Watson, M. S., McBride, K., and Davison, A. J. (1992). The DNA sequence of equine herpesvirus-1. Virology *189*, 304–316.

Tye, B. K., Nyman, P. O., Lehman, I. R., Hochhauser, S., and Weiss, B. (1977). Transient accumulation of Okazaki fragments as a result of uracil incorporation into nascent DNA. Proc. Natl. Acad. Sci. USA *74*, 154–157.

Upton, C., Stuart, D. T., and McFadden, G. (1993). Identification of a poxvirus gene encoding a uracil DNA glycosylase. Proc. Natl. Acad. Sci. USA *90*, 4518–4522.

Varshney, U., Hotcheon, T., and van de Sande, J. H. (1988). Sequence analysis, expression and conservation of *Escherichia coli* uracil DNA glycosylase and its gene (*ung*). J. Biol. Chem. *263*, 7776–7784.

Wittwer, C. U., and Krokan, H. E. (1985). Uracil-DNA glycosylase in HeLa S3 cells: interconvertibility of 50 and 20 kDa forms and similarity of the nuclear and mitochondrial form of the enzyme. Biochim. Biophys. Acta *832*, 308–318.

Worrad, D. M., and Caradonna, S. (1988). Identification of the coding sequence for herpes simplex virus uracil-DNA glycosylase. J. Gen. Virol. *62*, 4774–4777.

Yasui, A., Yajima, H., Kobayashi, T., Eker, A. P. M., and Oikawa, A. (1992). Mitochondrial repair by photolyase. Mutat. Res. *273*, 231–236.

Zell, R., and Fritz, H. J. (1987). DNA mismatch-repair in *Escherichia coli* counteracting the hydrolytic deamination of 5-methyl-cytosine residues. EMBO J. *6*, 1809–1815.

Zhang, K. Y. J. (1993). SQUASH: combining constraints for macromolecular phase refinement and extension. Acta Crystallogr. *D49*, 213–222.